Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other

Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP

Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)

Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud

Presenter: Lee Evans • Owner, LTS Computing LLC

Distributed Machine Learning Using OMOP

Presenter: Emily Pfaff • Research Assistant Professor, University of North Carolina at Chapel Hill

Clinical Registries and OHDSI

"OHDSI presents the opportunity to lower the cost of clinical registries by orders of magnitude through standardized EHR automation."

Clinical registries have a high human cost in chart abstraction

Trauma registry methodology: A survey of trauma registry custodians to determine current approaches

Gerard M. O'Reilly [a,b,*], Belinda Gabbe [a], Peter A. Cameron [a,b,c]

[a] *Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Commercial Rd, Melbourne, 3004, Australia*
[b] *Emergency and Trauma Centre, Alfred Health, Commercial Rd, Melbourne, Victoria 3004, Australia*
[c] *Emergency Services, Hamad Medical Corporation, Doha, Qatar*

0.5 FTE for every 200-300 patients.

Table 3
Human resources–Single hospital registries ($n = 40$).

Staffing type	Number of persons			
	0	1	2	>2
All (total) staff	0	2	4	34
Director/Head	14	25	1	0
Manager	18	21	1	0
Data manager	17	21	0	2
Database programmer	33	2	4	1
Database analyst	31	7	1	1
Trauma nurse coordinator	17	15	6	2
Data collector	12	2	4	22
Data entry clerk	23	9	2	6
ICD coder	26	4	0	10
AIS coder	21	4	2	13
Data analyst	28	9	2	1
Office administrator	29	11	0	0

Courtesy – Jon Duke, MD, Georgia Tech Research Institute

Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases

Patricia Biedermann, Rose Ong, Alexander Davydov, Alexandra Orlova, Philip Solovyev, Hong Sun, Graham Wetherill, Monika Brand & Eva-Maria Didden

BMC Medical Research Methodology **21**, Article number: 238 (2021) | Cite this article

Mapping registry data to the OMOP CDM facilitates more efficient collaborations between researchers and establishment of federated data networks

CURE ID and Virus COVID-19 Registry

- **CURE ID**
  - –Joint initiative between FDA, NIH/NCATS, Critical Path Institute
  - –Aims to identify repurposed drug candidates to treat infectious diseases
  - –Started as an online/app-based registry for clinicians to enter case reports
  - –COVID-19 expands mandate: automated extraction from electronic health records
- **Viral Infection & Respiratory Illness Universal Study (VIRUS) COVID-19 Registry**
  - –Launched by the Society of Critical Care Medicine within weeks of pandemic onset
  - –Rapidly described COVID-19 clinical course
  - –Global reach to 306 sites in 28 countries
  - –Started with manual data entry – hundreds of variables, ~4 hours per patient

CDRC
CURE DRUG REPURPOSING COLLABORATORY

CURE ID
Challenging Cases... Innovative Treatments

Goals for Cure ID

- Making OMOP accessible to community hospitals.
  - Create a feasibility checklist for clinical registries to adopt OMOP.
  - Lower the cost to ETL EHR data into the OMOP format.
  - Make it easier to deploy the OHDSI tool stack into a secure health system environment.

Goal 1: Clinical Registry OHDSI feasibility checklist

Question	Example	Answer
How many of the data elements correspond to core OMOP concepts?	Meds, Labs, Procedures, Conditions, Devices	Out of the Box
How many of data elements will need custom transformation scripts?	Vitals, Epic Flowsheets, Epic SmartForms	Minor Effort
How many of the data elements will need a custom concept?	Questions that are not in any standard lexicon.	Minor Effort
Will the OMOP data model needs to be extended to support the registry?	Imaging	Significant Effort
How many of the data elements will require NLP?	Symptoms at the time of admission.	Significant Effort

Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other

Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP

Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

- Matthew Robinson, MD is an infectious disease physician at JHU who also serves as faculty in Biomedical Informatics.

Adapting a manual clinical registry to OMOP

- Evaluate the feasibility of each variable for inclusion
- Weigh the challenge versus value of variables that require significant effort
- Consider the use case from the beginning
- Identify alternative means of data representation for complex variables
- Craft reproducible methods for variables requiring custom transformation

A Challenge - WHO Scale for COVID-19

- Not present as structured data in the EMR and infrequently documented in notes
- Identification of symptoms requires NLP
- Includes variables infrequently included in an OMOP ETL
  - –Respiratory support modality
    - Some modes of respiratory support lack standard terms
  - –Fraction of inspired oxygen

Patient State	Descriptor	Score
Uninfected	Uninfected; no viral RNA detected	0
Ambulatory mild disease	Asymptomatic; viral RNA detected	1
	Symptomatic; independent	2
	Symptomatic; assistance needed	3
Hospitalised: moderate disease	Hospitalised; no oxygen therapy*	4
	Hospitalised; oxygen by mask or nasal prongs	5
Hospitalised: severe diseases	Hospitalised; oxygen by NIV or high flow	6
	Intubation and mechanical ventilation, pO_2/FiO_2 ≥150 or SpO_2/FiO_2 ≥200	7
	Mechanical ventilation pO_2/FIO_2 <150 (SpO_2/FiO_2 <200) or vasopressors	8
	Mechanical ventilation pO_2/FiO_2 <150 and vasopressors, dialysis, or ECMO	9
Dead	Dead	10

Figure: **WHO clinical progression scale**
ECMO=extracorporeal membrane oxygenation. FiO_2=fraction of inspired oxygen. NIV=non-invasive ventilation. pO_2=partial pressure of oxygen. SpO_2=oxygen saturation. *If hospitalised for isolation only, record status as for ambulatory patient.

WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. *Lancet Infect Dis* 2020

Do we need this difficult variable – WHO Scale?

- Defines inclusion criteria and endpoints for COVID-19 trials
- Causal inference models must match similarly ill patients to isolate the impact of a therapeutic on clinical outcomes
- Lack of context from incomplete registry data contributes to misinterpretation

Patient A had an oxygen saturation of 95%

He received Drug X and survived

so I compared him to

Patient B had an oxygen saturation of 95%

He did not receive Drug X and died

Drug X saves lives!

https://www.nhlbi.nih.gov/health/ventilator

A Challenge - WHO Scale for COVID-19

- Feasibility
  - Out of the box to significant effort
- Use case and needs
  - Inpatient use requires only 4-10
  - Dropping gradations of mild disease obviates need for NLP
- Custom transformations
  - Respiratory support modality documented in flowsheets

Patient State	Descriptor	Score
Uninfected	Uninfected; no viral RNA detected	0
Ambulatory mild disease	Asymptomatic; viral RNA detected	1
	Symptomatic; independent	2
	Symptomatic; assistance needed	3
Hospitalised: moderate disease	Hospitalised; no oxygen therapy*	4
	Hospitalised; oxygen by mask or nasal prongs	5
Hospitalised: severe diseases	Hospitalised; oxygen by NIV or high flow	6
	Intubation and mechanical ventilation, pO_2/FiO_2 ≤150 or SpO_2/FiO_2 ≥200	7
	Mechanical ventilation pO_2/FIO_2 <150 (SpO_2/FiO_2 <200) or vasopressors	8
	Mechanical ventilation pO_2/FiO_2 <150 and vasopressors, dialysis, or ECMO	9
Dead	Dead	10

Figure: **WHO clinical progression scale**
ECMO=extracorporeal membrane oxygenation. FiO_2=fraction of inspired oxygen. NIV=non-invasive ventilation. pO_2=partial pressure of oxygen. SpO_2=oxygen saturation. *If hospitalised for isolation only, record status as for ambulatory patient.

WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. *Lancet Infect Dis* 2020

Wading through the flowsheets

- Flowsheets highly customized to individual EHR deployments
- Context of flowsheet entry matters
- Find flowsheet entries by name and by content
- Source to concept map must be customized to each site
- Protocols can improve reproducibility

Flowsheet entry name
LH Resp Dev
RN Oxygen Device
O2 Delivery H321

Flowsheet entry value
Room air
Ventilator
Nasal cannula
Room air

Goal 2: Lower the cost of ETL into the OMOP format

- Invest in the Perseus project
- An open source one stop ETL web tool
- A web-based GUI for ETL of OHDSI
  - Rabbit in the Hat
  - White Rabbit
  - Usagi
- Moved into OHDSI/Github Org on 2022-07-08
- Created a base configuration for common EHR's
  - Epic
  - Cerner (coming)

Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other

Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP

Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)

Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

- Zachary Wang is currently a Master's student at JHU in Biomedical Informatics and Data Science
- He is also one of our OHDSI Khierons!

Goal 3: Make it easier to deploy the OHDSI tool stack into Health Systems

- Improves data quality (DQD/Achilles)
- Helps Health System get greater value out of OHDSI
- Cure ID invested into the OHDSI BroadSea project
- Hosted with Microsoft Azure
- Docker Containers
- Security compliance

LTS Computing LLC

Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other

Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP

Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

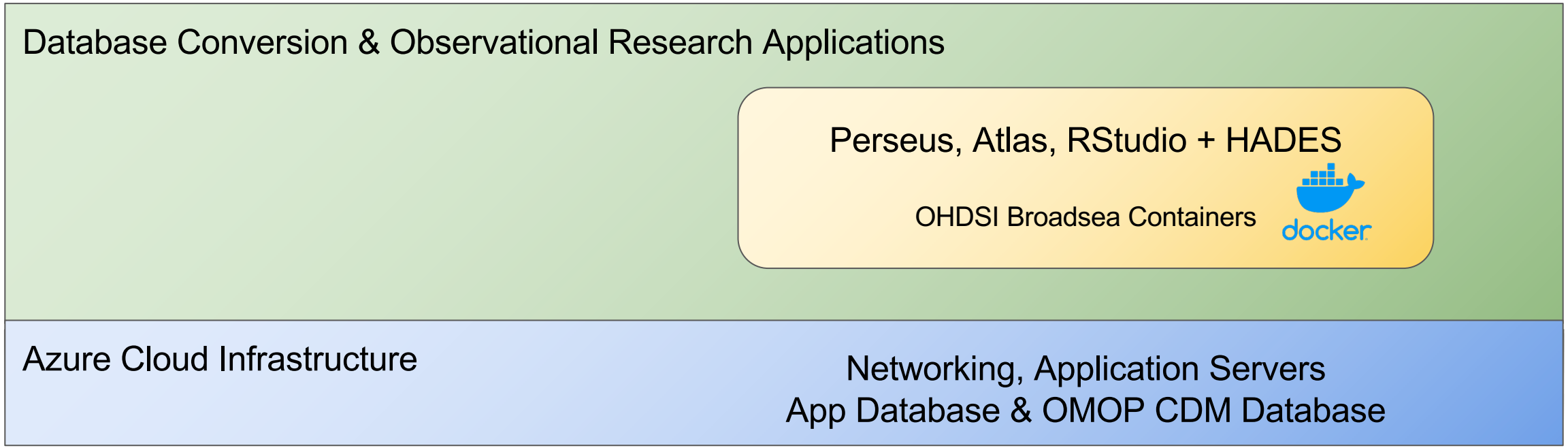
How to lower the ETL barrier going to OMOP with Perseus (Demonstration)

Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud

Presenter: Lee Evans • Owner, LTS Computing LLC

Solution Components Overview

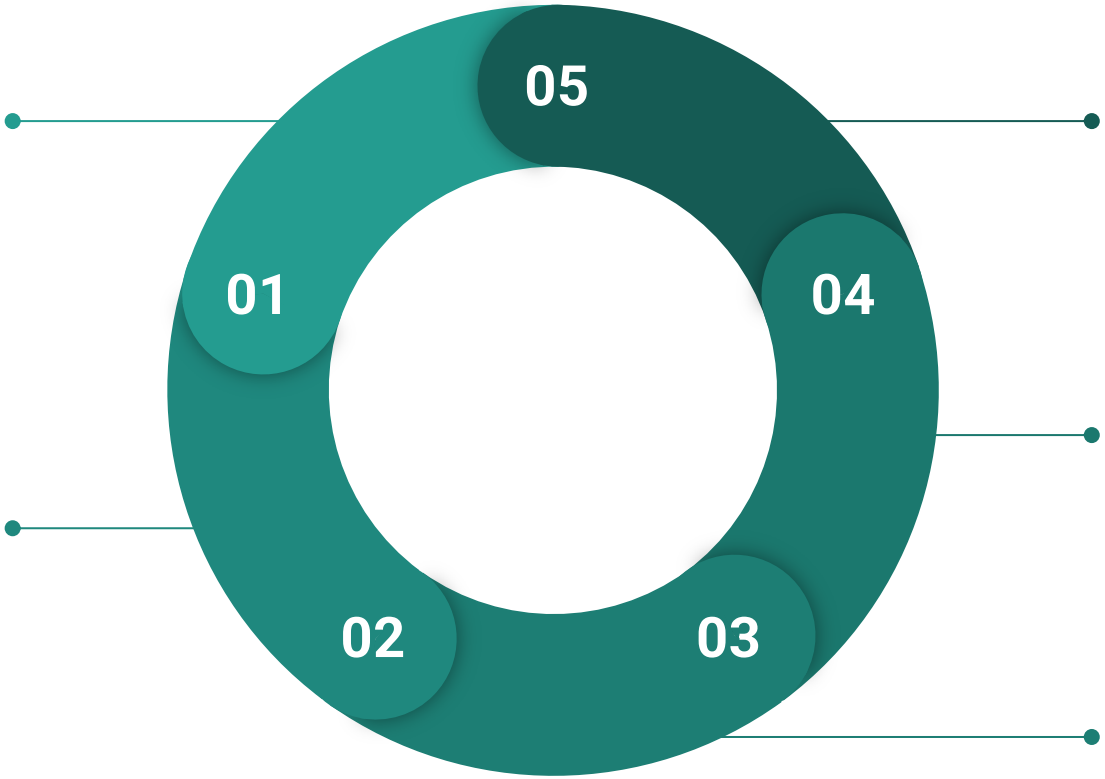


Perseus Open Source Extract Transform & Load (ETL)

Sergey Samus & his team

Scan Native Database
White Rabbit

Load base system mapping specification & customize
Rabbit In A Hat



Data Quality Assurance
Data Quality Dashboard

Convert Data to OMOP CDM

Map local codes to OMOP standard concepts
Usagi

OHDSI HADES Open Source observational analytics R packages

Population-level estimation

CohortMethod
New-user cohort studies using large-scale regression for propensity and outcome models.
Learn more…

SelfControlledCaseSeries
Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.
Learn more…

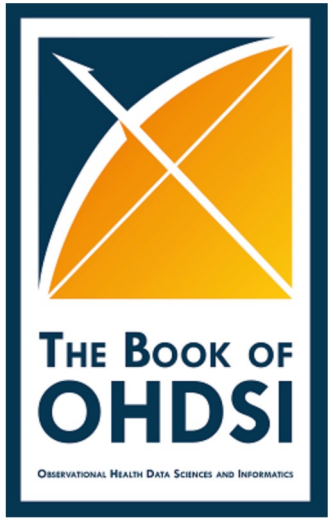
SelfControlledCohort
A self-controlled cohort design, where time preceding exposure is used as control.
Learn more…

EvidenceSynthesis
Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study.
Learn more…

Patient-level prediction

PatientLevelPrediction
Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.
Learn more…

EnsemblePatientLevelPrediction
Building and validating ensemble patient-level predictive models.
Learn more…



Cohort construction

Capr
Develop and manipulate complex cohort definitions in R
Learn more…

CirceR
An R wrapper for Circe, a library for creating cohort definitions, expressing them as JSON, SQL, or Markdown.
Learn more…

CohortGenerator
Instantiating cohorts in a database based on a set of cohort definitions.
Learn more…

PhenotypeLibrary
The OHDSI Phenotype Library: a collection of community-maintained pre-definined cohorts.
Learn more…

Evidence Quality

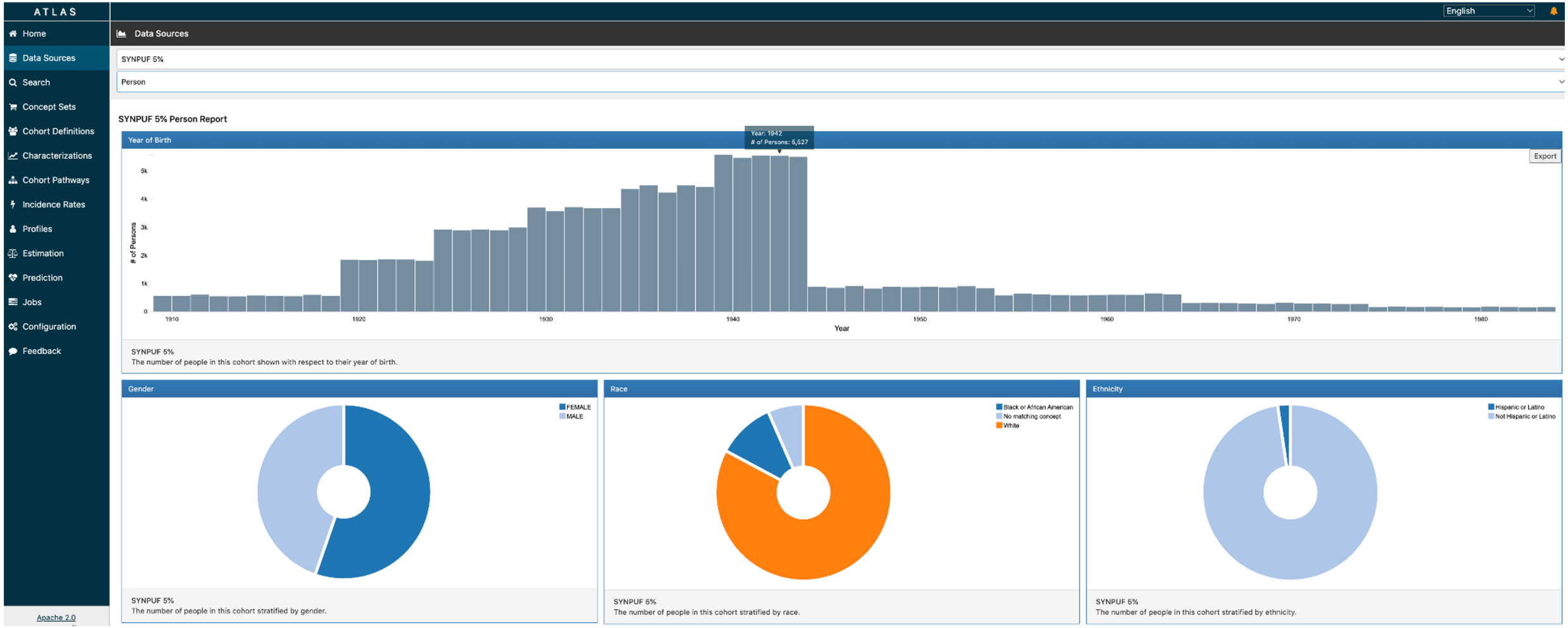
EmpiricalCalibration
Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.
Learn more…

MethodEvaluation
Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.
Learn more…

CohortDiagnostics
Generate a wide set of diagnostics to evaluate cohort definitions against databases in the CDM.
Learn more…

Open Source Book
https://ohdsi.github.io/TheBookOfOhdsi/

OHDSI ATLAS Open Source cohort designer



OHDSI Broadsea-Webtools Open Source

LTS Computing LLC

Docker Container

Web User Interface

3

Atlas Cohort Designer

Application Services

2

WebAPI Java App running in Apache Tomcat Server

App Libraries & Drivers

1

Java Runtime, Javascript, JDBC database drivers

“Zero to Atlas in ten minutes”

Install

git clone ohdsi/broadsea
docker pull ohdsi/broadsea-webtools
docker pull ohdsi/broadsea-eunomia

Run

docker-compose up -d

docker

OHDSI Broadsea 2.0 - FDA funded improvements

Security

- Open Source for transparency
- Docker images scanned for Common Vulnerabilities & Exposures (CVEs)
- Images cryptographically signed by OHDSI - Docker Engine can optionally verify image signature

Versioning & Meta-data

- Public Docker Hub repository retains older versions of images
- Reproducibility - e.g. a study may specify recommended Broadsea-Hades version
- Meta-data - additional provenance details included in image
- Automated build through GitHub actions (* soon)

More Broadsea Images

- Broadsea-Webtools - OHDSI Atlas
- Broadsea-Hades - RStudio Server + Hades R packages (replaces Broadsea-WebMethods)
- Broadsea-Eunomia - PostgreSQL database pre-loaded with Eunomia synthetic dataset
- Broadsea-Usagi - Perseus version of Usagi code mapping tool with web UI (* soon)

Azure Cloud Infrastructure Deployment

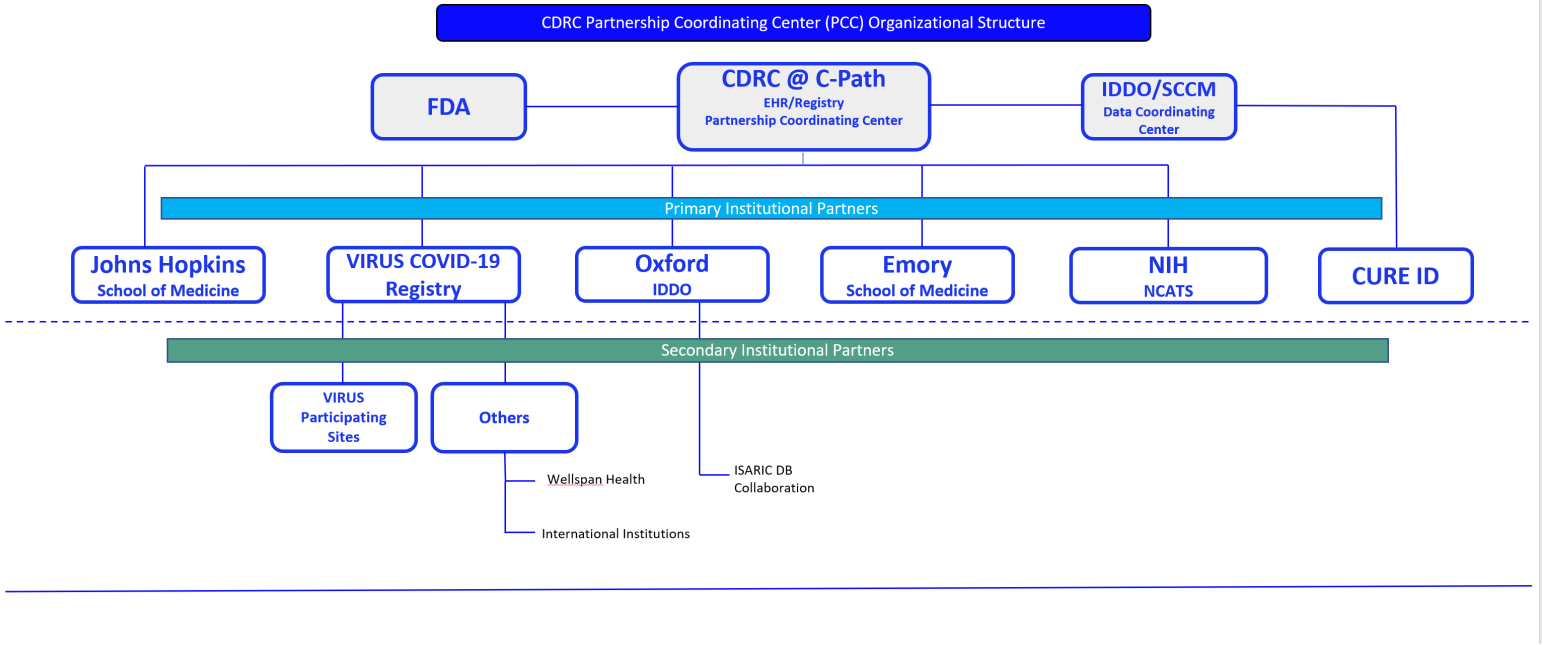
Microsoft Open Source Solution - Cory Stevenson & his team

Install Pipelines on Azure DevOps Service	Create Cloud Infrastructure	Install Application Software	Load OMOP Vocabulary Data
Clone **OHDSIOnAzure** GitHub repository Set Variable Groups to customize environment	Execute Terraform script to create cloud infrastructure	Launch OHDSI Broadsea Docker Containers on App Service	Download OMOP vocabulary data from OHDSI Athena website and load into database

Also, “OHDSIonAWS” open source solution (no Perseus yet) & potential future Google Cloud Platform solution.

CURE ID Pilot Results

- Pilot site: reduced ETL effort by 80% to 200 hours
- Recruiting sites to implement Perseus
  - Join VIRUS COVID-19 registry
  - Funding available
- Expanding beyond COVID
  - Meningitis
  - Sepsis
- More information:
  - CDRC@c-path.org



https://github.com/OHDSI/CureIdRegistry

Sept. 13: Clinical Registry Efforts in OHDSI

How clinical registries and OHDSI can benefit from each other

Presenter: Paul Nagy • Program Director for Graduate Training in Biomedical Informatics and Data Science, Deputy Director of the Johns Hopkins Medicine Technology Innovation Center

How to adapt a manual clinical registry to OMOP

Presenter: Matt Robinson • Assistant Professor, The Johns Hopkins University School of Medicine

How to lower the ETL barrier going to OMOP with Perseus (Demonstration)

Presenter: Zachary Wang • Graduate Student, Johns Hopkins (2022 Kheiron Cohort member)

Lowering the deployment burden with the cloud

Presenter: Lee Evans • Owner, LTS Computing LLC

Distributed Machine Learning Using OMOP

Presenter: Emily Pfaff • Research Assistant Professor, University of North Carolina at Chapel Hill